

# Improving Zero-Shot-Learning for German Particle Verbs by using Training-Space Restrictions and Local Scaling

Maximilian Köper, Sabine Schulte im Walde, Max Kisselew, Sebastian Padó

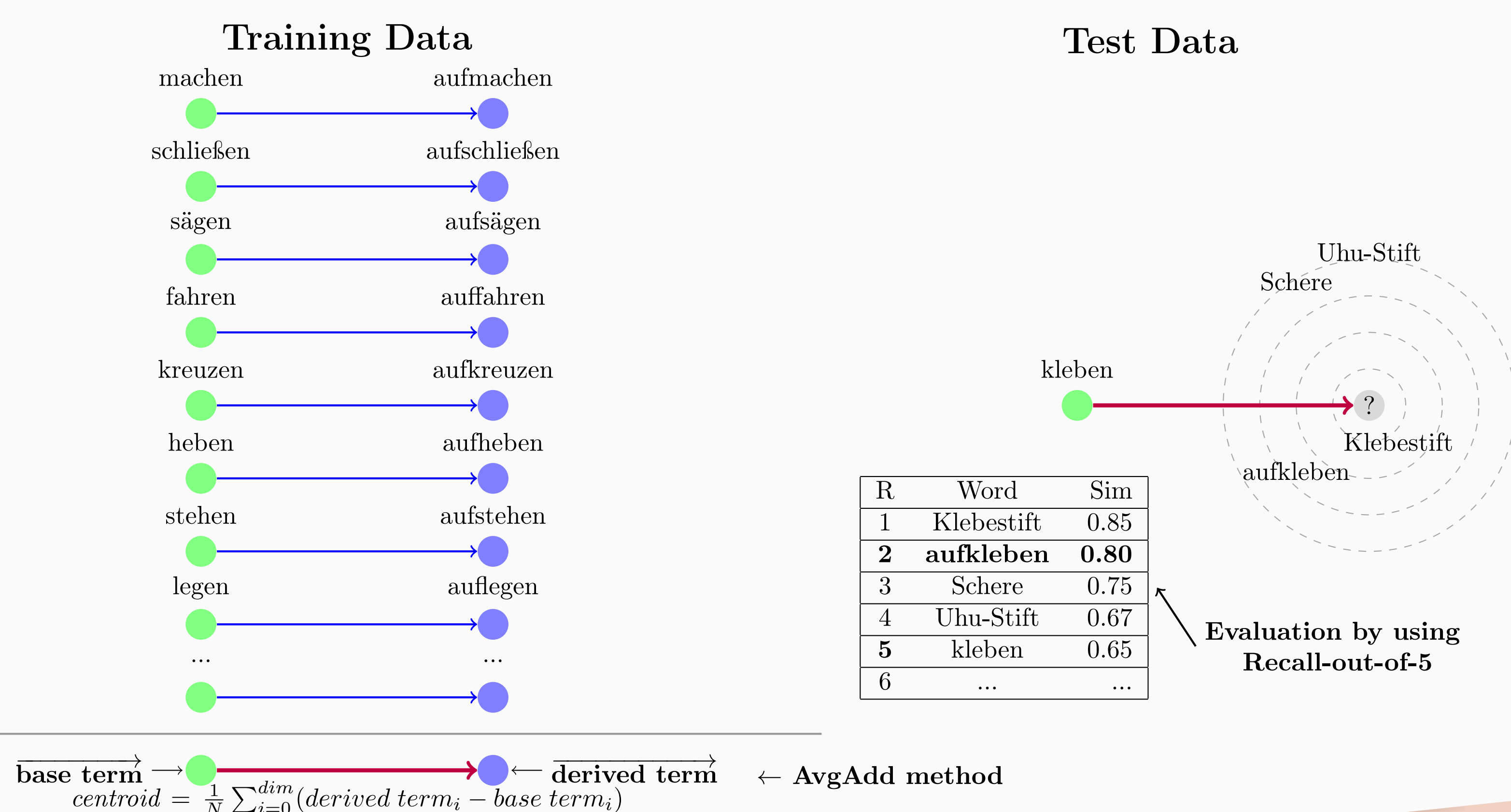
Institut für Maschinelle Sprachverarbeitung

Universität Stuttgart, Germany

{maximilian.koepfer,schulte,kisselmx,pado}@ims.uni-stuttgart.de



## Motivation



- Consider derivational patterns (e.g., *use* → *use + ful*) as the result of a compositional process, which combines base term and affix
- Learn a mapping function between base verbs (*lachen*) and particle verbs (*anlachen*)
- Main contribution: Experiments with various optimizations

## Dataset

- Experiments on **new dataset** (German particle verbs) +
- Existing derivational datasets from:

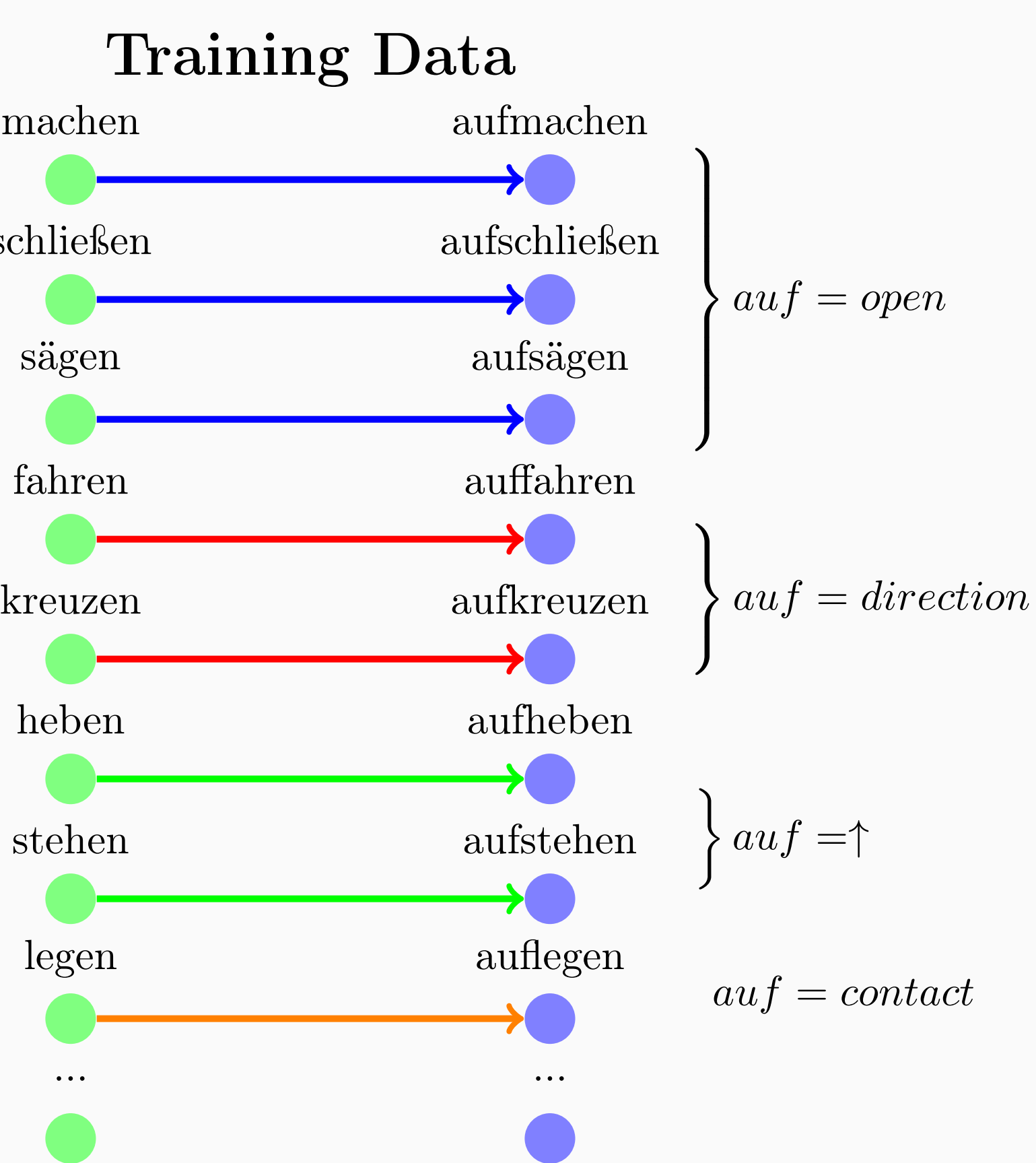
DE: Kisselew 2015 and EN: Lazaridou 2013

Affix	Example	Instances
auf-	nehmen - aufnehmen	171
ab-	setzen - absetzen	287
mit-	streiken - mitstreiken	216
ein-	laufen - einlaufen	185
zu-	drücken - zudrücken	50
an-	legen - anlegen	221
aus-	malen - ausmalen	280

**Data available!!**

[www.ims.uni-stuttgart.de/data/pv-deriv-dataset](http://www.ims.uni-stuttgart.de/data/pv-deriv-dataset)

## Restricting the Training Space (BestAdd)



- Motivation: particles are ambiguous; thus particle+verb derivations often undergo various meaning shifts

→ Training on a set of all particle verbs is counterproductive

→ Exemplar-inspired strategy: Restrict training data for a given base verb to the *k* base verbs with the highest cosine similarity

Test instance: *kleben* is most similar to *legen* (thus rely on **orange** training observation)

## Semantic Vector Space

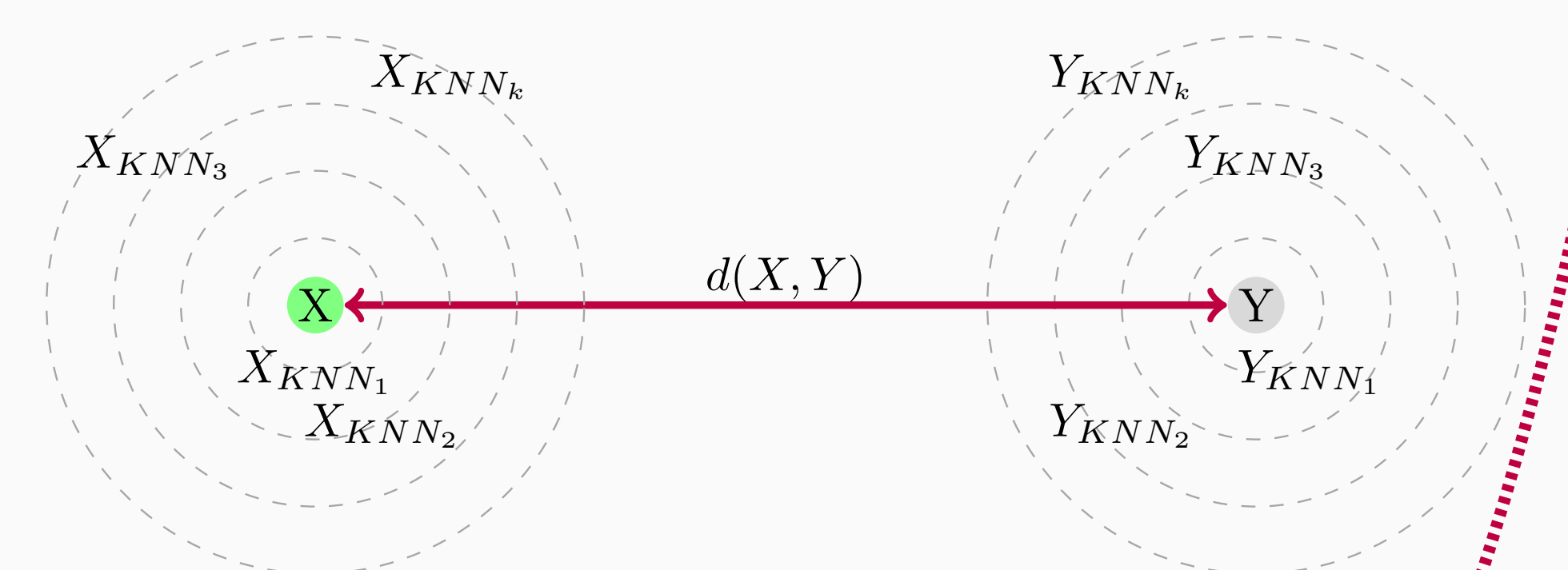
- Built on German (20B) and English (9B) COW 2012 corpora
- Use corpus lemma-POS pairs (*schlafen\_V*, *Haus\_NN*, ...)
- Symmetrical window of size 5 (only nouns, verbs, adjectives)
- 400 dimensions (svd, ppmi, frequency smoothing)
- 460 000 lemmas for German, 240 000 lemmas for English

## Local Scaling

- Improve Nearest-Neighbor search

$$LS(X, Y) = \frac{d_{XY}}{\sqrt{\mu_X \cdot \mu_Y}}$$

*LS* relies on the average distance  $\mu$  of *X* and *Y* to their *k* nearest neighbors



## Experimental Setup

- 10-fold cross-validation
- Baseline: predict base as derived term

## Results

Method	Particle Verbs (DE)		Kisselew (DE)		Lazaridou (EN)	
	Default	+LS <sub>15</sub>	Default	+LS <sub>15</sub>	Default	+LS <sub>15</sub>
Baseline	10.79%		16.08%		15.36%	
AvgAdd	11.82%	+1.28%	24.26%	+3.14%	24.19%	+2.95%
BestAdd <sub>1</sub>	10.22%	+1.19%	33.91%	+3.97%	27.32%	+1.87%
BestAdd <sub>3</sub>	<b>14.26%</b>	<b>+2.24%</b>	<b>38.50%</b>	<b>+4.17%</b>	37.06%	+1.40%
BestAdd <sub>5</sub>	14.44%	+1.97%	38.07%	+4.61%	<b>38.49%</b>	<b>+2.12%</b>

## Summary

- General improvement through training space restriction + Local Scaling
- Particle verbs remain challenging!

